DOCUMENT RESUME

ED 457 699 FL 026 935

AUTHOR	Fuentes, Alejandro Curado
TITLE	Representativeness and Significance Factors in ESP Texts.
PUB DATE	2000-00-00
NOTE	25p.; Published in Iberica 2:43-56.
PUB TYPE	Reports - Research (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	College Students; *Computational Linguistics; *English for Academic Purposes; *English (Second Language); Foreign Countries; *Grammar; Higher Education; Language Styles; *Lexicography; Limited English Speaking; Reading Skills; Second Language Instruction; Second Language Learning
IDENTIFIERS	Spain

ABSTRACT

The development of communicative approaches and strategies in specialized discourse has led to revising notions of representative and significant language. Particularly in work with academic genres, in science and technology settings, the need for determining these factors is growing. The application of specific language corpora becomes convenient. This paper seeks to specify the type of such corpus linguistic representativeness and significance sought in the case of teaching English to university computer science students. This study attempts to reveal the factors of representativeness and significance in the development of reading material, lexico-grammatical resources, and rhetorical strategies for an English for specific purposes (ESP) course. (Contains 46 references.) (KFT)

Reproductions supplied by EDRS are the best that can be made
from the original document.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Alejandro Curado

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

ED 457 699

Published in *Ibérica* 2: 43-56

REPRESENTATIVENESS AND SIGNIFICANCE FACTORS IN ESP TEXTS

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Alejandro Curado Fuentes
Universidad de Extremadura

The development of communicative approaches and strategies in specialized discourse has led to revising notions of representative and significant language. Particularly in the work with academic genres, in science and technology (EST) settings such as our own institution, the need for determining these factors is ever growing. The application of empirical resources such as specific language corpora, in fact, becomes rather convenient. In this paper, the aim is to specify the type of such corpus linguistic representativeness and significance sought in the case of teaching English to our groups of Computer Science students. In that scope, we present data and samples on which to base our suggestions and claims regarding the exploitation of textual material .

Key words: CORPUS, REGISTER, GENRE, LEXICO-GRAMMAR, FREQUENCY, RANGE, REPRESENTATIVENESS, SIGNIFICANCE

INTRODUCTION

The assessment of key lexis and grammar is conducted in planning courses of English for Academic Purposes (EAP) --under which we may locate EST

FL 026935

BEST COPY AVAILABLE

according to several authors (see, for instance, Jordan [1997]). The notion of register description underlined by Johansson (1975) is highly relevant in such a line of work. He refers to the need of designing computerized corpora in order to satisfy descriptive requirements of linguistic registers. For the teaching of English for Specific Purposes (ESP), in fact, as he states, frequency lists such as West's (1953) or Thorndike and Lorge's (1944),[1] "appear to be of limited use" in the design of course syllabi and material (Johansson, 1975:36). As conceived by Francis and Kucera (1964), in contrast, a large body of texts such as the Brown Corpus holds enough data for statistical inferences of lexical behavior. Improving or updating its category J --the register of scientific discourse--, for instance, may serve to provide linguistic and pedagogical awareness of change.

Underscoring this concern to depict linguistic variation, the analysis of register thus leads to investigating the level of words. As McCarthy states, "computer analysis is a very good way of getting at the vocabulary of a register" (McCarthy, 1990:64). Words, according to him, "acquire registerial appropriacy only in context" (103). The relationship is thus, in Carter's words, "dynamic" or "instantial": Words "make unique partnerships or combine or associate to produce meaning specific to that individual text" (Carter, 1997:177).

The notion of academic genre is also important in this respect. In contrast with that of register above, the former refers to textual distinctions or similarities 'on the basis of external criteria relating to the author's or speaker's purpose´(Biber, 1988:206). There may be two different genres, according to this view, and only one register of texts, such as biography and academic prose, which both have the narrative

linguistic form. Even more strictly, as Swales (1990:53) states, sub-genre distinction may be identified within a given genre, as administrative `good news´ letters vs. `bad news´ letters may prove -- Bhatia (1993) claims a similar analysis. The type of approach introduced makes the crucial distinction between genre and register dimensions of specialized languages or sub-languages. Not only Swales (1990) but Halliday and Hasan (1985) are implicitly present in this conception. The factors of coherence and cohesion are of prime importance, indeed, in marking out genre traits by focusing on words. Carter and McCarthy's joint assertion that lexis is " conditioned by genre" thus increases "the reader's predictive power and ability to create coherence" (Carter & McCarthy, 1997:205).[2]

As Aston asserts (1997:61), the syntagmatic level should thus be closely analyzed in relation with the paradigmatic plane. Within such a scope, computer corpora play a significant role, as these set the stage on which lexical items interact and perform bonds or associations. Setting the textual environment should be done judiciously in terms of gleaning language teaching priorities accordingly, e.g. grading suitable content and language which learners can profit from in terms of competence development. In this respect, the key for the opportune outset of such endeavors seems to be a well-founded selection of significant texts. For our purposes, this is primarily accomplished by demarcating the range of applicability on the EST curriculum. In other words, by defining the actual needs of language learning in academic settings as drawn from the evaluation of sources.

With these criteria in mind, we bring to bear measuring devices of textual

representativeness and lexical significance to our current ESP programs. In the following sections, the focus is placed on the description of these two factors as given by a corpus-based analysis of rhetorical and lexico-grammatical features in our own selection of academic material for Computer Science English (textbooks, technical reports and research articles).

METHODOLOGY

During the selection of texts for our pedagogical purposes, the enhancement of learning stages plays a crucial role. In Computer Science English, as James (1994) claims, it is very difficult to compile a corpus that is "representative of the language of Computer Science as a whole" (James, 1994:34). This scholar's aim is to work with "first-year Science and Technology texts, and to inform the construction of language teaching and learning materials in the light of these" (ibid., 34). Texts may thus be "differentiated according to subject matter, according to genre, or according to concept structure (information flow or topic type)" (ibid., 35). These factors are equally valued in our case; in such a manner, on developing an analysis of different genres in the register of academic writing for Computer Science, we regard the distinction of three learner levels.

In a first stage, it seems that excerpts, taken mainly from textbooks written by single authors are particularly suitable for first and second year university students, who should be able to identify key topics and concepts on the samples. The genre of technical reports may be added during this period of learning, since a large amount of examples of this type are published on the Internet, increasingly available to the readers. Subsequently, because of their reference and frequency of

use at higher stages, journal articles tend to occupy the advanced level of Computer Science English learning.[3]

The point of departure should therefore involve those academic approaches to Computer Science written exposition. The selection of what is meant as representative and significant language obeys, in this respect, two different parameters: The analysis of "linguistic variation" (Biber, 1988:13) across the text types, allowing for the perception of prevailing lexico-grammatical features, and, secondly, the examination of patterns and collocations of words, which aims at the description of sub-technical lexis.

In the first process of annotation mentioned, the key is to determine the representativeness factor in our ESP course by means of text analysis. Such a procedure entails taking each genre separately in order to compare functional patterns. The degree to which a chosen piece may thus embody specific linguistic traits may be measured through concordancing techniques. In contrast, for the development of significance observations, our Computer Science English corpus is taken as a whole and the data obtained framed as overall results. The words may then be recorded as either restricted or free lexical combinations.

Our work is carried out with three small sub-corpora corresponding to each aforementioned genre -- see complete references after Bibliography:

Sub-corpus A: One edited textbook dealing with the topic of Networks (Web design and Net description and history): 77,999 tokens; 6,673 types.

Sub-corpus B: Six technical reports about Hypertext technology (language programming and network structure): 86,361 tokens; 7,204 types.

Sub-corpus C: Four research articles on the subjects of Database and Graphics Systems (two texts) and Artificial Intelligence (two sources): 35,130 tokens; 2,905 types.

The first sample –sub-corpus A-- originally contained additional sections such as preface, introduction, bibliography and appendices. These have been removed in order to stick to the relevant content of the chapters. All the texts are downloadable from Internet locations.[4] Thus, compiling relevant texts can be hard in terms of search time and effort, while the issue of copyright permission is fairly easily coped with, as there tend to be fewer restrictions via hypertext documentation. Our choice of texts follows current concerns and priorities in course syllabi. Those are therefore based on the actual goals set out in compulsory subjects of Computer Science at our institution. The readings are suggested or recommended during the first, second, third and fourth years of studies. As Myers (1992: 9) and Conrad (1996: 302) assert, this variation among types of readings is necessary so that learners may widen both their knowledge and linguistic competence.

RESULTS

Biber (1988) suggests that what must be derived from such textual sets are co-occurring descriptive features of quantitative data. Thus, if certain characteristics are seen to "consistently co-occur, then it is reasonable to look for an underlying functional influence that encourages their use" (Biber, 1988:13).

This sort of analysis can thus be afforded, firstly, by rhetorical features, such as the use of purpose clauses and "we" statements across sub-corpora of specialized texts. Should some noticeable variation exist in the contrastive study, items may be organized accordingly, and the degree of representativeness for each text type (e.g. definition, description, exposition, etc) and genre (each sub-corpus) may be assessed in the light of resource applicability.

For the evaluation of purpose, indeed, the preposition "to" as a colligation of semi-technical words can work effectively in this sense.[5] The KWIC –Key Word In Context – function of the concordancer *WordSmith* (Scott, 1996) is highly convenient for such a display of data.[6]

Technical reports are often regarded as `semi-expert writing´ (Bergenholtz and Tarp, 1995:19) where the elaboration of objectives works as a main macrostructural device. The content is addressed to a semi-expert audience in an informational tenor, unlike textbooks, which tend to instruct instead (Myers, 1992:5). From this line of thinking, the notion of purpose should therefore develop in the form of end-constructions, such as the mentioned verbal company of the colligation "to". This is examined across sub-corpus B (technical reports), where the Collocation feature of the electronic concordancer enables the display of the following data:

WORD	L2	L1	R1	R2
ACCESS	2	104	13	11
HAVE	24	32	21	5
INFORMATION	11	5	10	39

INTERNET	7	6	6	59
ELECTRONIC	6	0	24	19
USE	5	0	48	13
CAN	8	0	0	0
USERS	4	15	10	10
PROVIDE	7	1	43	2

Figure 1: Most frequent content words collocating to the right (R) and left (L) of the node "to" --- L1, L2, R1, R2 refer to the positions occupied by the collocating words on both sides.
Numbers in each column = occurrences of the collocation at that slot.

Several examples are provided in figure 1, yet, as stated, the key feature to be pinpointed here is that the entry be a verb collocating immediately after the node. As a result, items like "access", "have", "use" and "provide" are likely candidates, co-occurring immediately after the node (Right 1 slots). As can be observed, with 43 occurrences, "provide" turns out to be the chosen form, since the others are either less frequent (e.g. "access" – 13 times) or more general English words ("have", "use"). This lemma –"provide"-- triggers a concordance (figure 2) that in fact reveals most uses of this form in end-clauses:

1	rked up with standardized tags in order to provide structure to the text (see
2	munity itself will build on those tools to provide special "campus
6	to document valid uses of tags in order to provide guidance for authors and
7	here to use links to outside documents to provide further reference inform
8	l treatment. Telemedicine has been used to provide medical consultation
10	ection to the Internet. In order to provide clarity to the reader, some s
11	main paths through the text, in order to provide a guidepost for those
12	hem terse. This article is just meant to provide pointers -- ---------
15	Level: 0 Function: Used to provide authorship information for HT

Figure 2: Concordance sample of the purpose colligation "to provide" in sub-corpus B.
Numbers = ranked positions of the lines according to text file sorting.

Figure 2 displays nine of the top 15 lines that exhibit the mentioned rhetorical function. The rest indicates a similar figure: 18 sentences express purpose by means of this semi-technical verb. From this scope, there seems to be a tendency

for "provide" to contain the prosodic information of stating a goal or objective in our sub-domain of Computer Science reports, as it can be deduced. Examining the cluster display of figure 3 below allows, in turn, the identification of the words that crop up as key collocates of "provide":

N	Cluster	Freq.
1	to provide a	8
2	in order to	6
3	order to provide	6
4	is to provide	4
5	provide the reader	4
6	the reader with	4
7	provide access to	3
8	used to provide	3

Figure 3: Clusters pointing to the semantic association of "to provide" with purpose.

These are phraseological occurrences that highlight the prevailing construction with "provide"; they constitute a fruitful source of linguistic study in Howarth's view (1996: 68-69), namely for the analysis of discourse and rhetorical functions across academic texts. A similar finding is gathered on examining other verbs, like "to use" and "to access", --for reasons of space, not shown here.

In terms of pedagogical implications, we may thus remark that in a first stage of reading textbook texts, purpose may be less explored. Yet, at a subsequent period, especially on dealing with the learners' approach to technical reports on specific computer issues, the rhetorical strategies tend to vary. Should the students have to cope with stating aim and cause, we glean in our body of texts that this is mainly done by working with the infinitive of verbs such as "provide".

In sub-corpus C -- research articles--, the frequency of "we" statements should equally facilitate hints on linguistic and teaching priorities. As in the reports

above, the contrast is carried out by checking whether variation is made noticeable. Thus, as figure 4 shows, the personal pronoun "we" in research articles is associated with procedural uses by means of sub-technical verbs like "define", "discuss", "use", "present", and "describe" – rather frequent and distributed across academic texts (McCarthy, 1990:50):

WORD	L1	R1	R2
DEFINE	0	10	3
DISCUSS	0	8	0
PRESENT	0	4	0
DESCRIBE	0	4	0
USE	0	4	4

Figure 4: Collocation chart excerpt of procedural verbs with the personal pronoun "we" in research articles.

In addition to the present simple patterns "we define" and "we use", seven examples of future statements are also included -- R2 in the rows of "define" (3 instances) and "use" (4). The present perfect is also employed significantly –not shown due to space demands. The data corresponds to the most frequent verbs accompanying "we" in sub-corpus C, whereas other forms include "see", "provide", "state", "market" and "find".

In a similar analysis across technical reports, in contrast, the results drawn may lead to infer the employment of the first person plural pronoun in more informal settings. This is mainly signalled by the co-occurrence of nouns with "have" (six instances) –figure 5-- and the present and past forms of "be" followed by gerund

expressions (12 times), as figure 5 indicates:[7]

WORD	L2	L1	R1	R2
ARE	0	0	9	1
OUR	2	0	0	2
WILL	0	0	8	0
HAVE	1	0	6	0
WERE	0	0	3	1

Figure 5: Patterns of auxiliary and delexical verbs with the personal pronoun "we" in technical reports.

These co-occurrences tend to mark out a less rigid procedural tone by indexical or delexicalized sign posting. Such an effect underlines a more interactive mode of discourse.

The sub-corpus of textbooks evokes a parallel attitude on the part of the writer in his / her relationship with the reader. The pattern of "we" + indexical verbs involves, in fact, a considerably large number of uses. In addition to the verbs in technical reports, others, like "would", "had", "could", "didn't", "know", and "want" also convey a sense of congeniality. Figure 6 gathers these appearances and computes the number of instances that are relevant from the perspective explored:

N	Word	Total	Left	Right
4	HAVE	28	3	25
5	WERE	27	5	22
6	HAD	26	3	23
7	ARE	21	4	17
13	WOULD	12	3	9
14	GET	11	2	9
16	COULD	10	1	9
20	DIDN'T	9	3	6
22	KNOW	9	4	5
24	OUR	9	4	5
25	WILL	9	0	9
67	WANT	5	0	5

Figure 6: Identification of auxiliary and indexical verbs collocating with "we" in textbooks. N = ranked position of the word in the Collocates chart.

The lexical traits, conforming the top 67 collocates of "we", are taken from the Collocation chart, organized according to frequency. In contrast with sub-corpus C, the words provide a distinct characterization of the handled textbook excerpts. In fact, given the data in figures 5 and 6 above, a hypothesis might be framed regarding "we" statements across technical reports and textbooks. Such a rhetorical feature is mainly used to express procedural statements. Indeed, it also seems to be often employed less rigidly, locating the speaker or writer in relation to the discourse and audience. In contrast, the more strictly academic or formal passages seem to favor a different approach: that of impersonal passive sentences. This is especially true in the case of technical reports, whereas there tends to be a combination of both "we" and passive uses in the case of journal articles.[8]

In this sense, anticipation of learning strategies and resources may be effected. The corpus serves as the medium eliciting the type of approach to be undertaken concerning teaching material and activity design. Biber et al.'s definition of representative corpus is thus a postulation to be followed: "The representativeness

of the corpus determines the kinds of research questions that can be addressed and the generalizability of the results of the research" (Biber et al., 1998:246).

For the second factor under inquiry –significance—, itemizing by means of subject-based lexical resources is a chief function, as mentioned above. These are entries bearing a significant semantic burden from a conceptual rung. In this respect, the vocabulary that highly co-occurs across all text types is of prime importance. Generally, these content items are both frequent and distributed across the texts. For instance, the term "information" occupies the 24th position (with 786 occurrences) in our corpus –see figure 7 below--. The same vocable is number 64 on the HKUST corpus of Computer Science English (James, 1994).

Observation of word behavior leads to the analysis of semi-technical lexical collocations. These prove reliable in the definition and characterization of concepts and processes, as noun phrases do in Engineering English (Varantola, 1984:30). The lemma "information" illustrates the type of lexis that provides the degree of both subject content and language significance. Frequency, in this respect, is not the only yardstick. As Pedersen (1997:65) states, specialized multi-word terms may keep relevant associations by appearing only twice in specific texts --as long as these are representative in given scientific-technical domains. Within such a scope, the Detailed Consistency Analysis –a function in WordSmith Tools--, which supplies frequency counts according to each genre, offers the occurrences of the word "information":

R	Lemma	RA	TR	TX

24	Information	15	618	153

Figure 7: Placement of the lemma "information" on the detailed consistency list of our corpus. R = ranking / RA = research articles / TR = technical reports / TX = textbooks. Numbers = frequencies.

The first row on the left indicates ranked position according to frequency and range across all texts and genres. The arrangement is determined by means of the distribution factor of the lexical item: how both often and evenly it appears throughout the corpus. In this sense, a word like "data", for instance, also occurs quite frequently; yet, its 109 instances in technical reports are offset by the mere three occurrences in research articles.

The remaining columns in figure 7 present the detailed frequencies in each given sub-corpus. As displayed, the three genres contain instances of the word uses. In order to check whether they, in fact, show significant associations, co-occurrences across texts must be examined. These are complementary in the sense that a given multi-word unit acquires significance in terms of its use across texts and genres. "Information", in fact, is identified more frequently in the technical report sub-corpus, but co-occurs significantly in other cases, as the items listed in figure 8 can attest. According to this table, the first set of lexical items corresponds to subject-based restricted collocates -- ® --whereas the second group includes freer combinations -- F --(cf. Howarth, 1996:68-69):

®
"Information services" (46), "library and information science" (30),
"network information" (24), "information systems" (23),
"information retrieval" (22), "information society" (13),
"information management" (10), "information resources" (10),
"information sources" (10), "information center" (8),
"information processing" (8) "information storage" (8),

"networking information" (8), "information providers" (7), "information studies" (7), "information tools" (7), "information flow" (5) "standard information" (6), "information science" (5), "information needs" (5)
F "electronic information" (24), "access to information" (22), "information technology" (13), "general information" (13), "information on the Internet" (12), "provides information" (10), "research on the information" (10), "information workers" (9), "information through" (8), "links to the information" (8), "stored information" (8), "information briefings" (6), "information held" (6), "information specialists" (6), "information and data" (5), "data and information" (5), "information available" (5), "information world" (5), "specific information" (5), "information community" (5), "facilitate information" (5), "any information" (3), "specific information" (2)

Figure 8: Lexical Collocations and clusters of the node "information" .
® = Restricted collocates / F = free combinations. Numbers = occurrences.

The distinction between types of lexical co-occurrences is thus based on the condition of their specific use within a limited number of texts or across topics. For instance, the expression "information available", a free association, is applied on a wider range of contexts than "information retrieval". In order to inspect these environments of use, the concordance is shown (figure 9). The former group tends to be uttered less technically than the latter:

1	I'd rather have the information available,
2	perceptions of electronic information available via the
3	load images, there is no information available. This
4	her when they made the information available? Do
5	having electronic information available is a
6	as made its product information available via the

1	maintaining to use networked information resources,
2	and Higgs' "Electronic Information Resources
3	aiming to use networked information resources,
4	information of how networked information resources
5	indexing and searching for information resources

6	internal and external information resources, a

Figure 9: Concordance samples of free and restricted combinations (first and second respectively).

Lines one and two on the top table correspond to the textbook, while three and four, to report # 2. Five and six, in turn, belong to report # 3 –see corpus references after bibliography. The list of restricted associates as exemplified by the use of "information resources", are, instead, identified almost exclusively in report # 3, describing virtual facilities on the Internet.

DISCUSSION

For our purposes of teaching Computer Science English as an academic register, the data gathered constitutes a minimal sample of semi-technical lexis and phraseology. At the Polytechnic setting in which we work, development and management of this linguistic knowledge should be applied according to each learning period. Thus, understanding the structure of definitions, descriptions and explanations in concepts and processes is mainly focused upon during the first year. At this point, students are made aware of multi-word constructions, but they are not likely to produce them until the intermediate / advanced levels, in which writing summaries or stating objectives of reports is a key task. The degree of significance that the lexico-grammatical chunks thus have increases through the courses. This is mainly due to their frequency and constancy across the reading curriculum.

Beginners generally tend to approach the text content by focusing on the syntagmatic level of words. This approach can neglect the overall discourse

structure, and important aspects such as textual coherence and cohesion may be left out. A common foe for the ESP instructor is truly embodied in such a conflict. As coordinating the two planes –word and context—is not an easy task, a unique answer to the problem is clearly beyond the scope of this paper. However, we may fulfill the objective set out in the introduction by offering a glimpse of how co-texts can function as a reduced reflection of broader subject contents.

From such a line of work, dealing with syntagmatic items can illustrate the graded sequence in the learner's acquisition process. This linguistic ability scale can be contrasted with H. Palmer's `ergonic system´, as defined by Howatt (1985:238): At an early stage, according to this view, the student is exposed to a limited set of words that amount to sentence units. These patterns then serve as primary matter samples for subsequent levels of learning. Through the `Direct Method´, in fact -- influenced by the behaviorist approach--, intermediate and advanced learners acquire the previous feedback to formulate secondary matter constructions. This scheme is also followed to some extent in the communicative approach, arising during the 1970s. In fact, according to Hornby during that period, as Howatt (1985:263) explains, verb patterns prove to work significantly as relevant clusters for sentence composition.

As an example, the sub-technical item "performance" may clarify this procedural concern, since the word occurs rather frequently in our corpus – 439 times. Its receptive decoding and encoding utilization can be located at varying periods of acquisition. The identification of the word use on reading textbook excerpts, for instance, introduces the pattern of the item by means of samples such as:

Web sites for performance spaces (Kitchen, Knitting Factory)
for National Research Initiatives, Performance Systems
The Federal High-Performance Computing Program," September 8,

Figure 10: Concordance of "performance" in beginner's textual material.

Taken as reference for subsequent activities, associations such as those marked out in figure 10 become labor-saving. Such compounds are chief aspects of language design in our ESP setting, and thus, propose primary matter with which patterns may be effected. Placing "performance" before "sytems" and "programs", or after "high" and "low" actually sets off as a key co-textual device.

Other instances of the node in technical reports provide contextual feedback in the sense of apportioning delimitation according to discourse functions. Realization and distinction of lexical uses become core elements by this judgement. The process of activating pattern writing receives great emphasis from this perspective. For example, on the construction of lines of the sort "the team was unable to develop and test performance measures to assess the utilization and impact of the network", compound items like "develop performance", "test performance" and "performance measures" interact with the rhetorical condition of purpose or goal ("to assess the utilization and impact of the network"). These, exploited receptively at the beginner's language plane, should be increasingly produced in academic tasks and exercises focusing on achievement of pattern awareness.

The end results sought, in this manner, are reflected by the abilities to understand and convey appropriate phrasing according to subject matter. Thus, for speaking,

for instance –usually rated by most students as their main concern in language learning--, delivering oral reports is a befitting academic micro-skill. During the performance of the task, the strategies of repetition and paraphrasing can be of great aid. Having been conveniently utilized before, these devices can be applied to resolve process statements in the presentation of graded information. This is denoted by students in the statement of clauses such as "our method provides very close performance to most of the JPEG modes" or "although the performance of the proposed method is not valid, our solution of performance measurement...". Our study thus attempts to reveal the factors of representativeness and significance in the development of reading material, lexico-grammatical resources and rhetorical strategies for our ESP course. In terms of exploiting competence, our corpus source is viewed as a potential element of negotiation. Items for such a meaning establishment are reliable discourse procedures, rhetorical features and lexical feedback. These are are highly valued in the work and support of EAP approaches to the Computer Science English class.

END NOTES

[1] According to Kennedy (1992), interest in delimiting lexis for teaching purposes led some scholars to develop lists of words for language learning in USA and Europe (Kennedy, 1992:335): Thorndike and Lorge came up with a 20 million word corpus that served to prepare a lemmatized list of 30,000 words in 1944. Theirs was the first to indicate frequency and range for each item, whereas West's *General Service List* (1953) is often taken as a reference source for lexical studies.

[2] Swales's work enables the distinction of text types in EAP and EST –e.g. research papers, reports, and other sources of task design. Halliday and Hasan

(1985: 61) state that "obligatory elements that define the genre to which a text belongs" (1985:61) are "text-specific" lexical relations, including collocations for grouping and defining words in texts (also cf. Hasan, 1984:183).

[3] Posteguillo (1997) finds a similar categorization of Computer Science English texts: mainly textbooks and research sources.

[4]Three useful web pages for the selection of readings can be found at the following addresses:
For textbooks,
http://www.utas.edu.au/docs/library/scitech/Electret.html
For reports, visit
http://www.research.digital.com/SRC/publications/src-rr.html
And for research articles,
http://www.lanl.gov/archive/cs

[5] Semi-technical words are highly important in the analysis of different genres, since these are words that occur most frequently across texts (Cowan, 1974: 390), embrace a wide range of contexts (Richards, 1974: 74), have a whole range of meanings (Herbert, 1965: 18) and should therefore be focused on in teaching EST (Inman, 1978: 246). Two comprehensive works on EAP and ESP –Jordan (1997: 152) and Dudley-Evans & St. Johns (1998: 100) – also give useful accounts of the importance of semi-technical words in specific academic environments.

[6] WordSmith Tools (1996) is mainly chosen on account of its useful features for the analysis of collocates and key words --functions like Clusters, Word list, and Consistency List, indeed quite helpful for our study.

[7] Indexical or delexical verbs are described as words that contain little content but are not considered function items (e.g."have", "get", "make", etc) (McCarthy, 1990: 51).

[8] For reasons of space, the concordance and collocates samples obtained in the analysis of passive sentences across the three sub-corpora are not shown. In general, technical reports contain the greatest number of passive statements (840 - -33.2%), while textbook texts show the lowest figure (12.3%). This agrees with the fact that textbooks include a large amount of active "we" sentences.

BIBLIOGRAPHY

Aston, G. (1997) "Enriching the Learning Environment: Corpora in ELT" in A. Wichmann et al. (eds.) *Teaching and Language Corpora*. London: Longman: 51-65.

Bergenholtz, H. & S. Tarp (1995) *Manual of Specialized Lexicography*. Amsterdam: John Benjamins.

Bhatia, V.K. (1993) *Analyzing Genre: Language Use in Professional Settings*. New York: Longman.

Biber, D. (1988) *Variation Across Speech and Writing*. Cambridge: Cambridge UP.

Biber, D. et al. (1998) *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge UP.

Carter, R. (1997) "Vocabulary, Cloze and Discourse: An Applied Linguistics View" in R. Carter & M. McCarthy (eds.) *Vocabulary and Language Teaching*. London: Longman: 161- 181.

Carter, R. & M. McCarthy (1997) "Lexis and Discourse: Vocabulary in Use" in R. Carter & M. McCarthy (eds.) *Vocabulary and Language Teaching*. London: Longman: 200-220.

Conrad, S. M. (1996) "Investigating Academic Texts with Corpus-Based Techniques: An example from Biology", *Linguistics and Education*, 8: 299-326.

Cowan, J.R. (1974) "Lexical and Syntactic Resources for the Description of EFL Reading Materials", *TESOL Quarterly*, 8: 389-399.

Dudley-Evans, T. & M.J. St. Johns (1998) *Developments in English for Specific Purposes*. Cambridge: Cambridge UP.

Farrell, P. (1990) *Vocabulary in ESP: A Lexical Analysis of the English of Electronics and a Study of Semi-Technical Vocabulary*. Dublin: Trinity College.

Francis, W.H. & H. Kucera (1964) *Manual of Information to Accompany A Standard Corpus of Present-day Edited American English, for Use with Digital Computers*. New York: Brown UP.

Halliday, M.A.K. & R. Hasan (1985) *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective*. Geelong: Deakin UP.

Hasan, R. (1984) "Coherence and Cohesive Harmony" in J. Flood (ed.) *Understanding Reading Comprehension*. Newark: International Reading Association: 181-219.

Herbert, A.J. (1965) *The Structure of Technical English*. London: Longman.

Howarth, P.A. (1996) *Phraseology in English Academic Writing*. Tübingen: Niemeyer.

Howatt, A.P.R. (1985) *A History of English Language Teaching*. Oxford: Oxford UP.

Inman, M. (1978) "Lexical Analysis of Scientific and Technical Prose" in M. T. Trimble et al. (eds.) *English for Specific Purposes: Science and Technology*. Oregon: Oregon State UP: 242-256.

James, G. (1994) *English in Computer Science: A Corpus-Based Lexical Analysis*. Hong Kong: Longman.

Johansson, S. (1975) *Some Aspects of the Vocabulary of Learned and Scientific English*. Gothenburg Studies in English 42.

Jordan, R.R. (1997) *English for Academic Purposes*. Cambridge: Cambridge UP.

Kennedy, G. (1992) "Preferred Ways of Putting Things with Implications for Language Teaching" in J. Svartvik (ed.) *Directions in Corpus Linguistics*. Berlin: Mouton de Gruyter: 335-373.

McCarthy, M. (1990) *Vocabulary*. Oxford: Oxford UP.

Myers, G.A. (1992) "Textbooks and the Sociology of Scientific Knowledge", *English for Specific Purposes*, 11: 3-17.

Pedersen, J. (1997) "The Identification and Selection of Collocations in Technical Dictionaries", *Applied Linguistics*, 19: 59-67.

Posteguillo, S. (1997) *Network English: English in the World of Computers*. Castellón: Publicacions de la Universitat Jaume I.

Richards, J.C. (1974) "Word Lists: Problems and Prospects", *RELC Journal*, 5: 69-84.

Scott. M. (1996) *WordSmith Tools*. Oxford: Oxford UP.

Swales, J.M. (1990) *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge UP.

Thorndike, E.L. & I. Lorge (1944) *The Teacher's Word Book of 30,000 Words*.

New York: Columbia University Teachers' College.

Varantola, K. (1984) *On Noun Phrase Structures in Engineering English*. Turku: University of Turku.

West, M. (1953) *A General Service List of English Words*. London: Longman.

CORPUS REFERENCES

Sub-corpus A:

Textbook # 1: Hauben, M. (1995) *Netizens: On the History and Impact of Usenet and the Internet.*

Sub-corpus B:

Report # 1: American Chemical Society Report # 3. (1995) "Environmental Science and Technology".

Report # 2: Birreli, A. et al. (1994) "Network Objects". Digital SRC Research Report # 115.

Report # 3: Chaiken, D. et al. (1998) "The Virtual Book". SRC Research Report # 157.

Report # 4: Crocca, W.T. & W.L. Anderson (1998) "Delivering Technology". Xerox Co.

Report # 5: Stephenson Von Tetzchner, J. (1998) "Frame 2 HTML Filter". Telecom.

Report # 6: Tilton, J.E. (1998) "Composing Good HTML".

Sub-corpus C:

Article # 1: Davis, E. (1999) "Order of Magnitude Comparisons of Distance", *Journal of Artificial Intelligence Research*, 10: 1-38.

Article # 2: Durr, C. & M. Chroback (1999) "Reconstructing h-v Convex Polyominoes from Orthogonal Projections", *Information Processing Letters*, 69: 283-289.

Article # 3: Hogg, T. (1999) "Solving Highly Constrained Search Problems with Quantum Computers", *Journal of Artificial Intelligence Research*, 10: 36-46.

Article # 4: O'Rourke, J. (1999) "Computational Geometry Column 33", *IJCGA*, 8, 3: 381-384.